



Generic GPCR residue numbers - aligning topology maps while minding the gaps

Isberg, Vignir; de Graaf, Chris; Bortolato, Andrea; Cherezov, Vadim; Katritch, Vsevolod; Marshall, Fiona H; Mordalski, Stefan; Pin, Jean-Philippe; Stevens, Raymond C; Vriend, Gerrit; Gloriam, David E

Published in:
Trends in Pharmacological Sciences

DOI:
[10.1016/j.tips.2014.11.001](https://doi.org/10.1016/j.tips.2014.11.001)

Publication date:
2015

Document version
Peer reviewed version

Document license:
[Unspecified](#)

Citation for published version (APA):
Isberg, V., de Graaf, C., Bortolato, A., Cherezov, V., Katritch, V., Marshall, F. H., Mordalski, S., Pin, J-P., Stevens, R. C., Vriend, G., & Gloriam, D. E. (2015). Generic GPCR residue numbers - aligning topology maps while minding the gaps. *Trends in Pharmacological Sciences*, 36(1), 22-31.
<https://doi.org/10.1016/j.tips.2014.11.001>

Generic GPCR residue numbers – aligning topology maps while minding the gaps

Vignir Isberg¹, Chris de Graaf², Andrea Bortolato³, Vadim Cherezov⁴, Vsevolod Katritch⁵, Fiona H. Marshall³, Stefan Mordalski^{6,7}, Jean-Philippe Pin^{8,9}, Raymond C. Stevens^{4,5}, Gerrit Vriend¹⁰, and David E. Gloriam¹

¹ Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

² Division of Medicinal Chemistry, Faculty of Sciences, Amsterdam Institute for Molecules, Medicines and Systems, VU University Amsterdam, The Netherlands

³ Heptares Therapeutics Ltd, Welwyn Garden City, UK

⁴ The Bridge@USC, Department of Chemistry, University of Southern California, Los Angeles, CA 90089 USA

⁵ The Bridge@USC, Department of Biological Sciences, University of Southern California, Los Angeles, CA 90089 USA

⁶ Department of Medicinal Chemistry, Institute of Pharmacology, Polish Academy of Sciences, Krakow, Poland

⁷ Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Krakow, Poland

⁸ Institute of Functional Genomics, Centre National de la Recherche Scientifique (CNRS) Unité Mixte de Recherche 5203, Universities Montpellier, Montpellier, France

⁹ Institut National de la Santé et de la Recherche Médicale (INSERM) Unité 661, Montpellier, France

¹⁰ Centre for Molecular and Biomolecular Informatics (CMBI), Radboudumc, Nijmegen, The Netherlands

Generic residue numbers facilitate comparisons of, for example, mutational effects, ligand interactions, and structural motifs. The numbering scheme by Ballesteros and Weinstein for residues within the class A GPCRs (G protein-coupled receptors) has more than 1100 citations, and the recent crystal structures for classes B, C, and F now call for a community consensus in residue numbering within and across these classes. Furthermore, the structural era has uncovered helix bulges and constrictions that offset the generic residue numbers. The use of generic residue numbers depends on convenient access by pharmacologists, chemists, and structural biologists. We review the generic residue numbering schemes for each GPCR class, as well as a complementary structure-based scheme, and provide illustrative examples and GPCR database (GPCRDB) web tools to number any receptor sequence or structure.

New era in GPCR research

G protein-coupled receptors (GPCRs) constitute the largest family of human cell surface receptors [1]. They respond endogenously to ions, neurotransmitters, lipids, carbohydrates, nucleotides, amino acids, peptides and proteins; and also sense light, pain, tastes, and odors [2]. Their abundance in human physiological systems, as well as their accessibility and druggability, have made them a major drug target

family – ~30% of the marketed drugs act on GPCRs [3]. The GPCRs are typically classified into the classes A–F [4], or according to the alternative GRAFS (glutamate, rhodopsin, adhesion, frizzled/taste2, secretin) system for the human receptors [5] (Table 1). In recent years, the number of GPCR crystal structures has grown exponentially and are now available for all of the human classes (A–C, F), except the Taste type 2 receptors. These structures have revealed common conformational changes during receptor activation, allosteric modulation by ions, lipids, cholesterol, and water; as well as interactions with the G protein [6–8].

Generic residue numbers - maps to navigate GPCR topology

All GPCRs share a structural core of seven transmembrane (7TM) helices, making up the machinery for signal transduction across the cell membrane. The 7TM domain contains or is part of the binding site of class A and B1 receptor ligands, and serves as a site for allosteric modulation of class B2, C, and F GPCRs [9,10]. So far, 109 GPCRs have been drugged [11], the vast majority with ligands binding within the TM region [12]. The conserved 7TM scaffold allows for the alignment of sequences or structures to identify the corresponding residues, which are indexed with a generic residue number. Such generic residue numbers allow for comparison of, for example, mutant effects, ligand interactions, and structural features across receptor subtypes, species orthologs, or receptor subfamilies. For example, such residue numbers have been used to define the shared ligand-accessible residue positions within the TM bundle [13]. These conserved positions are common placeholders, while receptor sequences vary and so compose the unique mosaic responsible for ligand affinity and selectivity.

Corresponding author: Gloriam, D.E. (david.gloriam@sund.ku.dk).

Keywords: G protein-coupled receptor; sequence alignments; mutational effects; ligand binding; structural motifs.

0165-6147/

© 2014 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tips.2014.11.001>

Table 1. Human GPCR classification and crystal structures^a

Class [1]	GRAFS [2] Family	Note	# Human receptors [3]	Crystal Structures (7TM domain)
A	Rhodopsin	Includes ~400 olfactory receptors [4]	282 (+ 400 olfactory)	2000: rhodopsin [5], 2007: β_2 -adrenoceptor; and today: 23 receptors in total [6]
B	Secretin	Also referred to as B1	15	2013: Corticotropin-releasing factor receptor 1 [7] and Glucagon receptor [8]
	Adhesion	Also referred to as B2	33	–
C	Glutamate	Includes Taste type 1 receptors	22	2014: Metabotropic glutamate receptor 1 [76] and 5 [77]
F	Frizzled		11	2013: Smoothed [78]
O (Other)	Taste2	Re-classified as own family [79], although originally grouped with Frizzled [2]	24	–

^aClasses D and E do not exist in human, and are fungal mating pheromone receptors and cAMP receptors, respectively.

Class A GPCR residue numbering

The Ballesteros–Weinstein numbering scheme [14] is based on the presence of highly conserved residues in each of the seven transmembrane (TM) helices. It consists of two numbers where the first denotes the helix, 1–7, and the second the residue position relative to the most-conserved residue, defined as number 50. For example, 5.42 denotes a residue located in TM5, eight residues before the most-conserved residue, P5.50. The residue numbers can be counted directly within the receptor protein sequence (alignment); however, the reference residues are not conserved in all receptors and the numbering is therefore not always straightforward (class A conservation: N1.50: 98%, D2.50: 90%, R3.50: 95%, W4.50: 97%, P5.50: 78%, P6.50: 99%, P7.50: 88% [15]).

Alternative class A numbering schemes have been presented by Oliveira [16], Baldwin [17,18], and Schwartz [19,20]. They have a common basis in enumerating residue positions from the helix extracellular ends, aiming to assign residues located at the same depth in the membrane with the same numbers, for example 3.16 and 6.16 (this reverses the TM2, TM4, and TM6 sequences). However, none of the schemes, which use different starting points and numbers, succeeds because GPCR crystal structures have uncovered extensive variations in the length and inclination of TM helices. The alternative schemes differ in format: Oliveira numbers (the oldest numbering scheme) omit the dot separator to make the numbers computationally more accessible, and Baldwin and Schwartz helix numbers are denoted with roman numerals, I–VII.

Class B, C, and F GPCR residue numbering

Class B, C, and F schemes have been established using the same procedure as the class A Ballesteros–Weinstein system, but use unique reference positions (X.50) such that the residue numbers can be counted directly within the receptor protein sequence (alignment). The class B GPCR Wootten [21] scheme is based on the B1/secretin subclass, but the reference residues are the most conserved also for five of the B2/adhesion receptor helices and the remaining two, TM3–4, still display high conservation (E3.50 58% and W4.50 42%) [15]. This scheme was used in the publications of the crystal structures of both the glucagon receptor [22] and corticotropin-releasing factor receptor 1 (CRF₁)

[23]. The class C GPCR Pin [24] numbering was used in the publication of the metabotropic glutamate receptor 5 (mGlu₅) crystal structure [25]. The class F GPCR Wang scheme was introduced in the recent publication of Smoothed (SMO) crystal structures [26]. In humans, this is a small class with only 11 members and, in cases where a helix has more than one fully conserved position, the one structurally closest to the class A Ballesteros–Weinstein was used as the reference position. Because all schemes use identical formatting, it has been suggested to append the class name (A–F) where clarification is needed, for example, 3.50b for class B Wootten numbers [27].

Cross-class GPCR residue numbering

The low sequence conservation between the GPCR classes has hitherto hindered (correct) sequence alignments, although some inter-class receptor modeling studies correctly aligned the majority of the seven helices (e.g., [28–30]). The structural conservation is higher and the recent crystallographic data have opened up for structure-based sequence alignments from class A to B [22,23,27], C [25,31], and F [32,33]. Some helices display large inter-class lateral deviations or different bending but, because adjacent helices are often translated in the same direction, structural multi-residue motifs with a shared functional mechanism are often conserved across the classes. The published cross-class residue comparisons have utilized the Ballesteros–Weinstein numbers, and where necessary included a class-specific number, for example Y7.53a.57b. Furthermore, reference cross-class alignments, based on the available crystal structures, are available in GPCRDB (<http://www.gpcrdb.org>; see below). Table 2 shows the alignment of the class specific Ballesteros–Weinstein numbers based on structural alignment of crystal structures of representative receptors from class A (bRho), B (Glucagon receptor), C (mGlu₁), and F (SMO).

Example 1: class A/B common receptor activation motif in TM7

A Tyr residue Y^{7.53a.57b}, conserved in both class A (Y7.53) and class B (Y7.57) GPCRs, has been proposed to play an important role in the activation of both receptor families (Figure 1) [34]. In the Glucagon receptor (class B GPCR) crystal structure [22] Y400^{7.57b} forms hydrogen bonds with

Table 2. Alignment of the class-specific generic residue numbers based on structural alignment of crystal structures of representative receptors from class A (bovine rhodopsin, bRho), B (glucagon receptor), C (metabotropic glutamate receptor 1, mGlu₁), and F (Smoothed, SMO). Class-specific reference residue positions (X.50) for each of the seven TM helices are given in bold font.

Helix	Class A (bRho)	Class B (GCCR)	Class C (mGluR1)	Class F (SMO)
TM1	G51 ^{1.46a}	S152 ^{1.50b}	G603 ^{1.50c}	T241 ^{1.43f}
	N55 ^{1.50a}	L156 ^{1.54b}	T607 ^{1.54c}	T245 ^{1.47f}
	T58 ^{1.53a}	A159 ^{1.57b}	V610 ^{1.57b}	T248 ^{1.50f}
TM2	L76 ^{2.43a}	H177 ^{2.50b}	C631 ^{1.39c}	L267 ^{2.42f}
	D83 ^{2.50a}	F184 ^{2.57b}	I638 ^{1.46c}	F274 ^{2.47f}
	L84 ^{2.51a}	V185 ^{2.58b}	F639 ^{1.47c}	F275 ^{2.50f}
	F88 ^{2.55a}	S189 ^{2.62b}	Y642 ^{2.50c}	S278 ^{2.53f}
TM3	L131 ^{3.46a}	E245 ^{3.50b}	K678 ^{3.50c}	A327 ^{3.46f}
	R135 ^{3.50a}	L249 ^{3.54b}	I354 ^{3.54c}	W331 ^{3.50f}
TM4	W161 ^{4.50a}	W272 ^{4.50b}	I714 ^{4.40c}	W365 ^{4.50f}
	P171 ^{4.60a}	W282 ^{4.60b}	L724 ^{4.50c}	I375 ^{4.60f}
TM5	V211 ^{5.45x46a}	P310 ^{5.42}	Y759 ^{5.46c}	P407 ^{5.50f}
	P215 ^{5.50a}	A314 ^{5.46b}	L763 ^{5.50c}	V411 ^{5.54f}
	I219 ^{5.54a}	N318 ^{5.50b}	C767 ^{3.54c}	G415 ^{5.58f}
TM6	L262 ^{6.45a}	G359 ^{6.50b}	C795 ^{6.47c}	L464 ^{5.45f}
	W265 ^{6.48a}	E362 ^{6.53b}	W798 ^{6.50c}	F467 ^{5.47f}
	P267 ^{6.50a}	V364 ^{6.55b}	A800 ^{6.52c}	C469 ^{5.49f}
	Y268 ^{6.51a}	F365 ^{6.56b}	F801 ^{6.53c}	H470 ^{5.50f}
TM7	A299 ^{7.46a}	G393 ^{7.50b}	V823 ^{7.40c}	F526 ^{7.46f}
	P303 ^{7.50a}	A397 ^{7.54}	L827 ^{7.44c}	I530 ^{7.50f}
	—	—	P833 ^{7.50c}	V536 ^{7.56f}

the conserved T351^{6.42b} and E245^{3.50b} residues [27] in a conformation that in class A GPCRs is linked to activation and interaction with the G protein [6,35]. In CRF₁, the thermostabilizing mutation Y363^{7.57b}A contributes to the shift in the conformation of the receptor towards an inactive state [23].

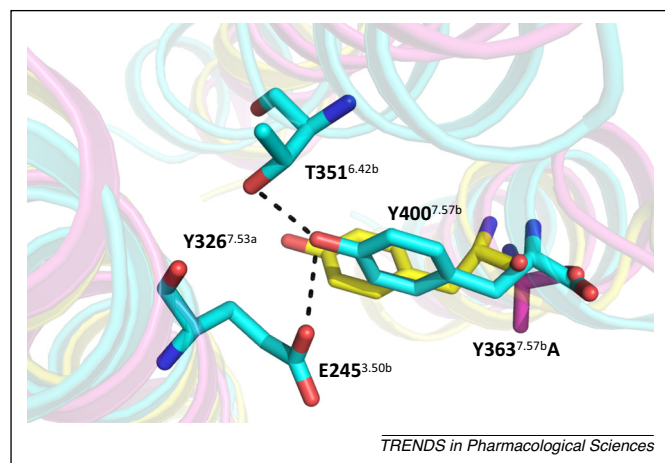


Figure 1. Class A/B common receptor-activation motif in transmembrane (TM) helix TM7. Intracellular view of structural superposition of constrictor-releasing factor receptor 1 (CRF₁) (magenta, PDB: 4K5Y), the glucagon receptor (cyan, PDB: 4L6R), and the β_2 -adrenoceptor (yellow, PDB: 3SN6). The conserved Tyr residue Y7^{5.7b} and in CRF₁, the thermostabilizing mutation Y363^{7.57b}A are shown in stick representation. In the glucagon receptor the polar interactions of Y400^{7.57b} with T351^{6.42b} and E245^{3.50b} residues are represented as broken lines.

Example 2: class A/C unique features of conserved residues in TM6 involved in ligand binding and receptor activation

In both classes A and C, TM6 contains a highly conserved tryptophan at position 6.48a.50c. This residue occupies an equivalent position in both classes, but in the structures of the metabotropic glutamate receptors mGlu₁ [31] and mGlu₅ [25] TM6 has moved laterally away from TM7 compared to class A. In contrast with class A, this movement allows W^{6.50c} to create a hydrogen bond to the backbone carbonyl of 5.44c, bridging TM6 to TM5. Despite this conformational difference, the aromatic residues W^{6.48a.50c} and F^{6.51a.53c} play a role in the ligand binding in both classes (allosteric ligands for class C), as shown in the class A [6,8] and C [25,31] crystal structures (Figure 2). Their role in ligand binding and potentially in receptor activation is supported by mutation studies for several class A aminergic [36,37], chemokine [38], adenosine [39], and lipid [40] receptors, as well as class C receptors mGlu₁ [41], mGlu₅ [42,43], and CASR (calcium-sensing receptor) [44,45].

Example 3: class A/B/C/F common hydrophobic core

In the helical bundle, TM3 represents a structural and functional hub [6] being centrally placed with contacts to all other helices except TM1. TM3 plays a central role in activation [6], which involves the upward movement of TM3 along its axis and a rotation of TM6 [7]. Comparisons of inactive and activated class A receptor structures have led to the proposal of a hydrophobic mechanism hindering activation that consists of the rearrangement of a core of highly conserved hydrophobic residues in 3.43a together with V/I/L6.40a, V/I/L6.41a, and F6.44a [7]. This core is well conserved in all human GPCRs: all classes A, B, C, and

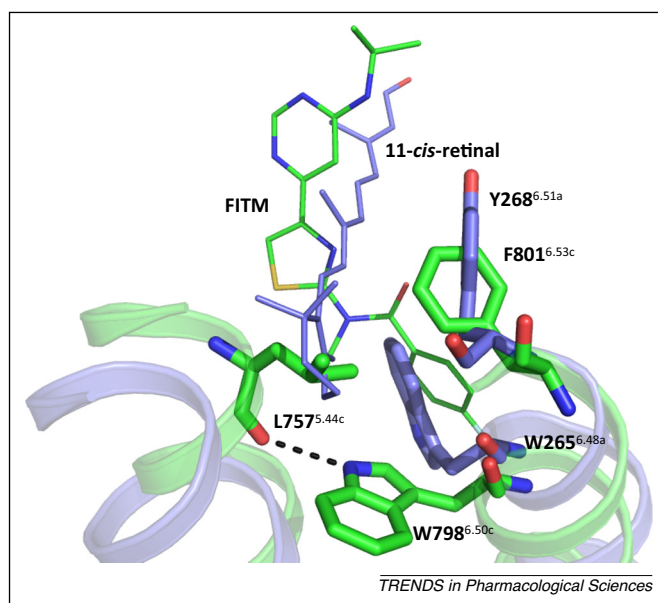


Figure 2. Class A/C unique features of conserved residues in TM6 involved in ligand binding and receptor activation. Side view of rhodopsin (purple, PDB: 2X72) and the metabotropic glutamate receptor mGlu₁ (green, PDB: 4OR2). 11-*cis*-retinal and FITM are shown in line representation. L^{5.44c} and the aromatic residues W^{6.48a.50c} and Y/F^{6.51a.53c} are shown in stick representation. In mGlu₁, the polar interaction between W^{6.50c} and L^{5.44c} is represented as a broken line. Abbreviation: FITM, 4-fluoro-N-[4-[6-(isopropylamino)pyrimidin-4-yl]-1,3-thiazol-2-yl]-N-methylbenzamide.

F have a highly conserved hydrophobic residue in position 3.43a.47b.47c.43f (mainly L, M, L, and W; respectively) that forms contacts with at least two hydrophobic residues in the equivalent TM6 positions 6.40a.45b.42c.39f, 6.41a.46b.43c.40f, and 6.44a.49b.46c.43f (recently described in Bortolato *et al.*) [15,33]. Furthermore, in all classes except C the other side of TM3 has hydrophobic contact with a highly conserved W4.50abf (same index in all classes) on TM4 [15,33].

Extracellular loop 2 residue numbering

Sequence analysis shows that there is a large diversity in the lengths and compositions of the N- and C-termini as well as in the extracellular and intracellular loops connecting the TM helices of GPCRs [46]. Nevertheless, for the extracellular loop 2 (EL2) a similar residue-numbering scheme has been applied [41,47] in which EL2 residues are labeled 45.X, indicating the location between TM4 and TM5 ("45") [47]. The reference position (X.50) is a conserved cysteine forming a disulfide bridge with a TM3 residue C3.25a.29b.29c.25f, which is fully (100%) conserved in the classes B, C, and F, and 88% conserved in class A [15,47]. The second extracellular loops of many GPCRs have approximately the same number of residues upstream and downstream of C45.50 [47]. However, the loop regions of GPCRs are structurally less conserved than the TM helices [6], and EL2 is no exception. Thus, in comparisons across receptor subfamilies the structural conservation is limited to the backbone of 45.50–45.52. This short stretch is still of high importance because the 45.52 sidechain is often directed into the 7TM binding pocket [47]. Longer loop modeling has to be reserved to receptors where there is a template with the same length in TM4–C45.50 and C45.50–TM5, respectively, and guided by experimental restraints. Moreover, although recent GPCR Dock challenges show that although the positions of specific EL2 residues close to C45.50 can be correctly predicted, computational modeling of complete EL2 structures is still highly challenging [48–50].

Mind the gap – introducing a novel complementary crystal structure-based GPCR residue numbering

Class A GPCRs have long been known to contain helix kinks in TM5–7 that are induced by highly conserved proline residues, which are unable to form backbone hydrogen bonds [51]. Upon receptor activation, they act as hinges that allow the helices to tilt inwards to tighten the ligand cavity and outwards to widen the G protein-binding pocket [7]. This does not affect the generic residue numbers because the best corresponding alignment of residues is obtained knowing that receptor structures are dynamic and the same helix in two receptors can in most cases tilt to adopt a similar position. Consequently, the GPCR 7TM was assumed to be uniformly shaped, leading to the long-standing principle to never insert gaps within sequence alignments of the TM helices.

As we entered the structural era, we discovered that the GPCR 7TM frequently contains distortions of another type: bulges and constrictions that are local to one helix turn, but

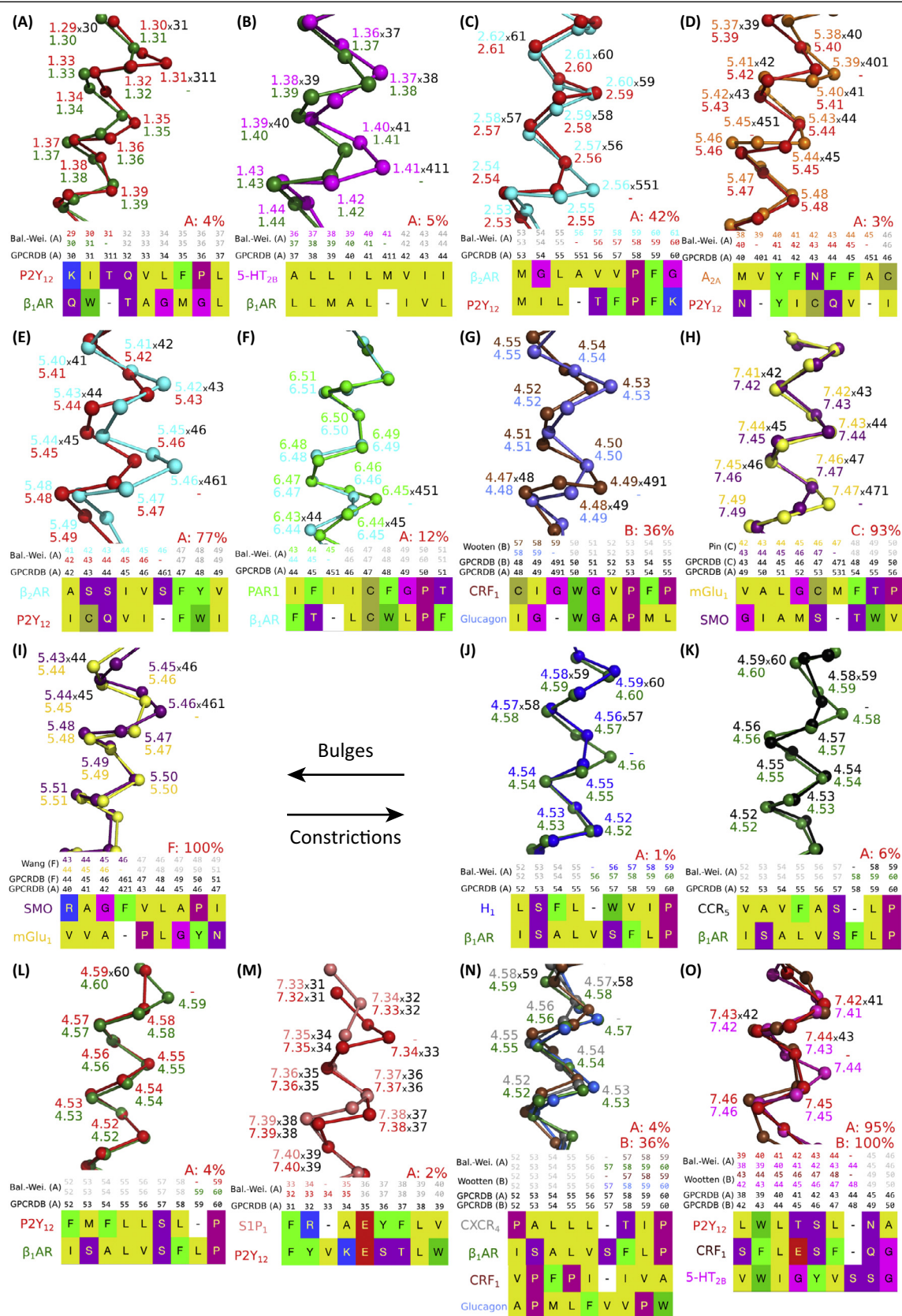
offset the generic numbers of all following residues in the helix. This is because there is one additional residue in a bulged (π -helical), and one absent residue in a constricted (3_{10} -helical) helical turn, respectively. This causes a single unmatched position, in other words a gap in structure and sequence compared to the corresponding undistorted helix. To date, more than 110 crystal structures of 27 unique GPCRs, have uncovered nine bulges (Figures 3A–I) and six constrictions (Figures 3J–O) in TM1, TM2, TM4, TM5, TM6, and TM7 [52]. These range all classes A, B, C and F; and two constrictions are shared by two classes (Figures 3N–O). Some helical distortions affect only a specific receptor subtype (Figure 3J), whereas others are shared by the majority (Figures 3E, H, and O) or all (Figures 3I and O) members of a GPCR class. Of note, the highly conserved proline residues P6.50x50 and P7.50x50 are associated with kinks in all available class A GPCR structures, but entail only few (12%, Figure 3F) or no observed bulges, respectively, in other words, they lack the backbone hydrogen bond but have no amino acid insertion.

GPCRDB has implemented a complementary structure-based scheme that corrects for bulges and constrictions. The scheme is based on pairwise superposition of inactive receptor structures using the 'WHAT IF' program [53]. Each helix is analyzed separately, and a bulge or constriction is defined by comparison to an undistorted reference. The single bulge residue that protrudes the furthest (Figures 3A–I) is assigned the same number as the preceding residue followed by a 1, for example, a bulge after residue 46 is given the number 461. The position lacking in a constriction (Figures 3J–O) is simply skipped in the residue numbering. Uncertain cases are put on a 'waiting' list of potential bulges/constrictions that require additional crystallographic evidence. GPCRDB numbers are assigned to all distortions, even those that seem to be shared by all members of a GPCR class. This is for two reasons. First, it allows cross-class structure/sequence alignments, and second, it future-proofs the residue numbers by avoiding updates should another class-member with an undistorted helix be crystallized.

To distinguish the GPCRDB scheme, it uses a unique separator x (e.g., 5x46) to denote that it is based on experimental structures, which are predominantly X-ray structures although NMR structures are also taken into account when available. For clarity, it is recommended to use the combination of the sequence- and structure-based numbers, for example 2.58x57, unless the comparison only includes crystallized receptors.

Example 4: TM2 bulge affects chemokine and opioid receptor binding site residues

Although most GPCR crystal structures contain an alpha bulge in TM2 [52] at position 2.55x551, chemokine receptors and opioid receptors contain a S/T^{2.56}XP^{2.58} sequence motif that stabilizes a different helical conformation. Site-directed mutagenesis data probing the TM2–TM3 interface [54] and receptor–ligand interactions in chemokine receptors [38,55] were used to successfully predict the TM2 conformation of CXCR4 (chemokine (C–X–C motif) receptor 4) and interactions between W94^{2.60} and D97^{2.63} and



TRENDS in Pharmacological Sciences

Figure 3. Bulges and constrictions in GPCR class A, B, C and F crystal structures. Bulged (A–I) and constricted (J–O) helical turns have one additional and lacking residue position, respectively, when superposed to an undistorted helix reference. The structural and sequence alignments of equivalent residues therefore contain a gap, which offsets all following generic residue numbers (gray). The GPCRDB generic residue numbers (black) are structure-based and account for bulges and constrictions by numbering an extra residue as the preceding followed by a 1, for example 1.40x411, and by skipping a lacking position. To facilitate cross-class comparisons, their alignments have been assigned GPCRDB numbers for all classes. The percentages to the lower right of the structures represent the conservation/frequency of the bulge or (Figure legend continued on the bottom of the next page.)

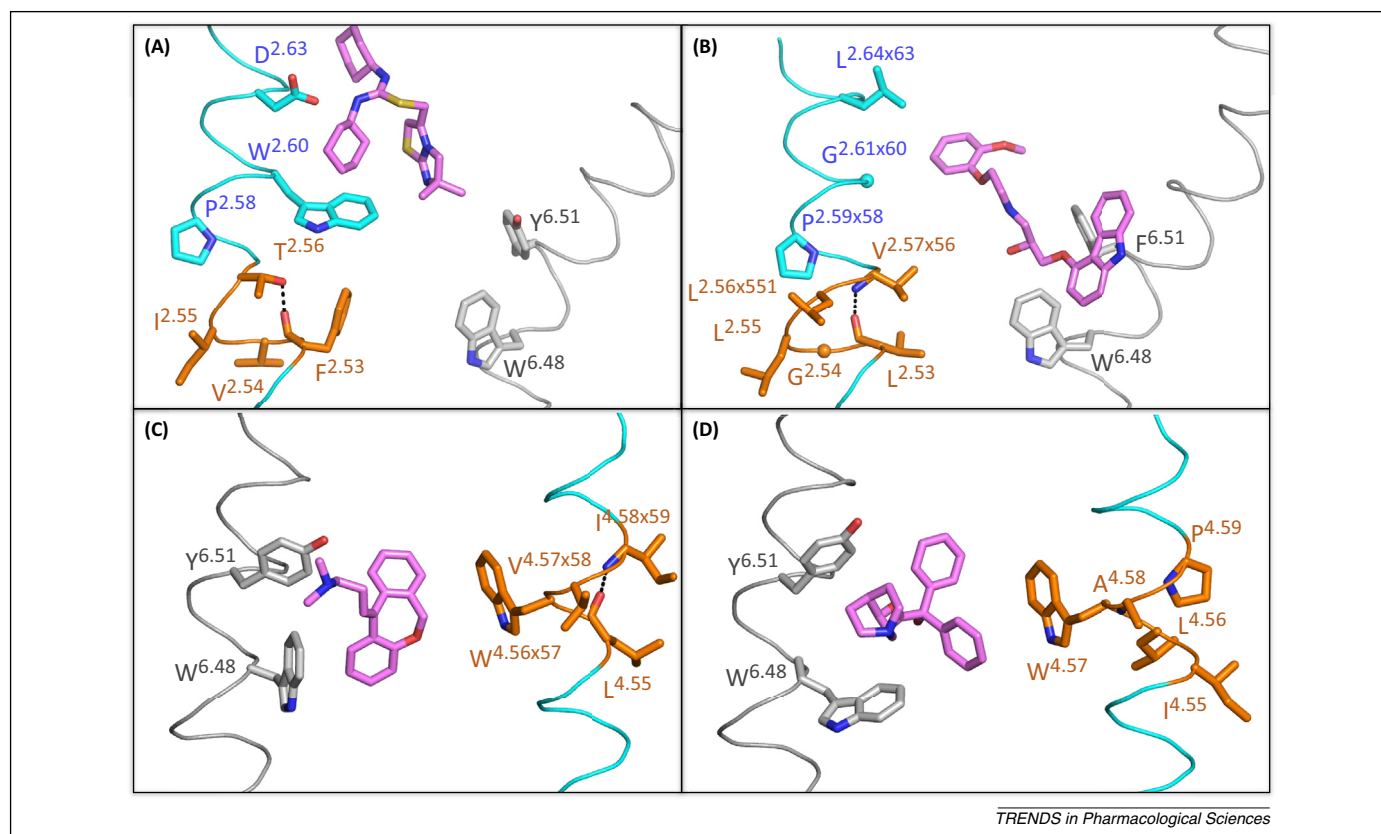


Figure 4. TM bulges and constrictions orient different residue numbers in the ligand binding site in GPCR crystal structures. Comparison of the T^{2.56}X^{P2.58} stabilized region (orange) in TM2 of the CXCR4 chemokine receptor that orients W^{2.60} and D^{2.63} into the 1T1t binding site [56] (A), and the alpha-bulge in TM2 of the turkey β_1 -adrenoceptor that orients G^{2.61x60} and L^{2.64x63} into the carvedilol binding site [62] (B). Comparison of the constricted region (orange) in TM4 of the histamine H₁ receptor [64] that orients W^{4.57} into the 3-quinuclidinyl-benzilate binding site (D). Ligand carbon atoms are colored magenta. Residues W^{6.48} and Y/F^{6.51} in TM6 are shown as reference (note that TM6 in panels A–B is rotated $\sim 90^\circ$ relative to TM6 in panels C–D).

the co-crystallized 1T1t ligand prior to the CXCR4–1T1t crystal structure [56] in the GPCR Dock 2010 challenge [49,57] (Figure 4A). CXCR4 models based on an ungapped sequence alignment between CXCR4 and the crystal structure templates available at the time of this GPCR modeling competition (rhodopsin, adrenoceptors β_1 and β_2 , and the adenosine receptor A_{2A}R) incorrectly oriented W94^{2.60} and D97^{2.63} towards the membrane layer instead of towards the ligand-binding site [49,57]. Residues at positions 2.60 and 2.63 interact with co-crystallized ligands in CCR5 (chemokine C–C motif receptor) (W^{2.60}, Y^{2.63}), the κ receptor (kappa 1 opioid receptor) (Q115^{2.60}, V118^{2.63}) [58], and NOP receptor (opiate receptor like-1 receptor) (Q107^{2.60}, D110^{2.63}) [59] crystal structures. Moreover, mutagenesis studies indicate that 2.63 (K108 and N129, respectively) plays a role in the ligand selectivity of δ and μ opioid receptors [55,60]. These residues structurally align with residues L101^{2.64x63} and H93^{2.64x63} in adrenoceptors β_1 (Figure 4B) and β_2 , respectively, two examples of GPCRs that contain a bulge in TM2. L101^{2.64x63} interacts with carmeterol, dobutamine, and carvedilol (Figure 4B) in β_1 -adrenoceptor crystal structures [61,62], while BI-167107 interacts with H93^{2.64x63}, and FAUC50 covalently

binds the H93^{2.64x63}C cysteine mutant in β_2 -adrenoceptor crystal structures [35,63].

Example 5: TM4 constriction affects the alignment of aminergic receptor binding site residues

The TM4 of the aminergic histamine H₁ receptor crystal structure [64] is constricted, directing W158^{4.56x57} (an important residue for H₁R ligand binding based on mutation studies [65]) towards the aromatic ligand-binding pocket (Figure 4C). The aminergic muscarinic acetylcholine M₂ and M₃ receptor crystal structures [66,67] are not constricted in TM4, and the structurally equivalent position of the conserved tryptophan is 4.57. This again lines the ligand-binding site and makes hydrophobic/aromatic contacts with the co-crystallized ligand, in accordance with mutation studies of W400^{4.57} in M₂ [68] (Figure 4D). Furthermore, combined mutagenesis and protein–ligand modeling studies to explain ligand selectivity for histamine H₃ and H₄ receptors [69] and H₄R species variants [70] indicate that the residues at position 4.57 (Y167^{4.57} in H₃R and N147^{4.57} in H₄R) are directed towards the ligand-binding pocket. This suggests that the TM4 of H₃R and H₄R are undistorted, making the constriction unique for the H₁R subtype.

constriction within the class, as seen in the GPCRDB alignments [15]. Adenosine receptor A_{2A}R has two bulges in TM5 (D) giving a double offset, for example, 5.39 versus 5x41. Spheres represent the alpha carbon of each residue. Receptors: 5-HT_{2B} (PDB: 4IB4), A_{2A}R (PDB: 4EIY), β_1 -adrenoceptor (PDB: 4AMJ), β_2 -adrenoceptor (PDB: 2RH1), CRF₁ (PDB: 4K5Y), H₁ (PDB: 3RZE), P2Y₁₂ (PDB: 4NTJ), S1P₁ (PDB: 3V2Y), SMO (PDB: 4JKV), and mGlu₁ (PDB: 4OR2).

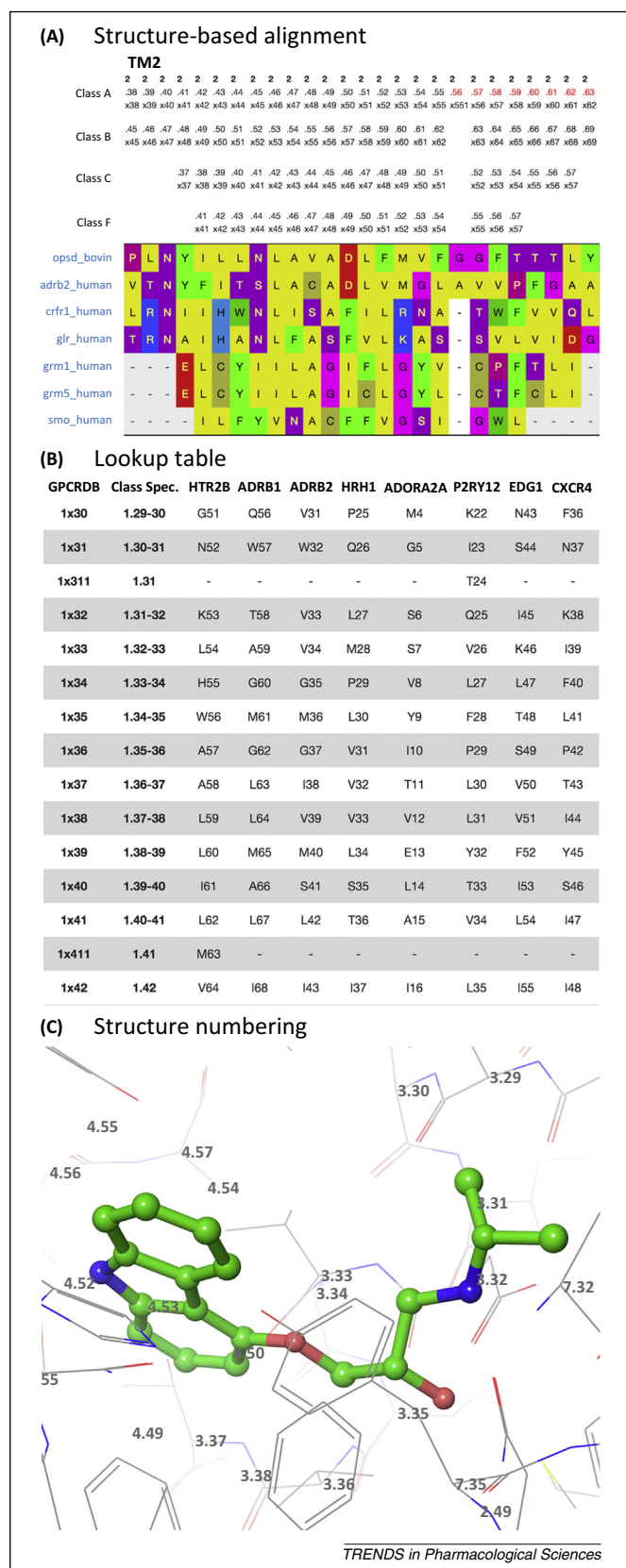


Figure 5. GPCRDB residue numbering tools. GPCRDB offers a suite of residue numbering tools. **A)** Sequence alignments are gapped to account for bulges and constrictions and list both the structure-based (GPCRDB) and sequence-based residue numbers, for each class. The example, obtained from the structure browser tool, shows the TM2 alignment for two crystallized class A–C receptors and one class F GPCR. **B)** Lookup tables show generic and receptor-specific residue numbers and can be used to compare the equivalent receptor residues. The example covers TM1 for the class A receptors included in the case stories. **C)**

Numbering made easy – residue-numbering web tools at GPCRDB

GPCRDB has been a major community resource for more than 20 years [15,71–73]. It contains reference data including the largest available collection of receptor mutants, crystal structures, 3D structure models in the inactive and active states, and sequence alignments for all species in UniProt. Recently, GPCRDB was equipped with a new suite of interactive web browser tools and diagrams; these include phylogenetic trees, sequence motif search (e.g., for a binding site), and receptor sequence plots (snake and helix box plots). Notably, for all of the above, correct alignment of residues in sequence and/or structure is crucial.

For pharmacologists, chemists, structural biologists, and others to use generic residue numbers, a more convenient way is needed than the manual generation of sequence alignments and assignment of residue numbers. Consequently, GPCRDB provides reference alignments and a series of residue-numbering tools. The sequence alignments are based on the available crystal structures and are gapped to account for bulges and constrictions (Figure 5A). GPCRDB supports alignments of all classes displaying both the sequence-based Ballesteros-Weinstein (A), Wootten (B), Pin (C), and Wang (F) numbering, as well as structure-based GPCRDB residue numbers. The alignments are followed by a consensus sequence and statistics on residue/property conservation, and can be downloaded for further analysis. A predefined set in the receptor selection allows alignment or lookup (below) of only the crystallized receptors.

Lookup tables show side-by-side listing of receptor-specific and generic residue numbers (Figure 5B). The tables can list multiple receptors, such as subtypes or species orthologs, to facilitate the inference of for example mutagenesis data or ligand interactions observed in crystal structure complexes. GPCR structures can be browsed or uploaded in Protein Data Bank (PDB) format to assign up to two generic numbering schemes to the TM residues (Figure 5C). The numbers are stored in the PDB file by replacing the B-factor of the C-alpha and carbonyl carbons and scripts are presented for visualization using PyMOL, Maestro, or MOE (Molecular Operating Environment). Finally, to allow compatibility and reinterpretation of published literature, the Utopia PDF reader [74] can be downloaded to annotate receptor residues with information from GPCRDB. A guide to generic GPCR residue numbering is presented in Box 1.

Potential exceptions and future directions

The sequence- and structure-based numbering systems share some limitations. Generic residues can only be assigned to receptor regions with a conserved structural fold, in other words the 7TM domain, whereas the termini and loops can only be compared, in the best case, within receptor subfamilies. Both systems also depend on sequence homology to produce sequence alignments, although this dependency decreases for the GPCRDB

Structures can be browsed or uploaded in PDB format to assign up to two generic numbering schemes to the transmembrane residues. The example shows the carazolol binding site of the β_2 -adrenoceptor (PDB: 2RH1).

Box 1. Generic GPCR residue numbering – how to

Sequence-based residue numbers are defined by class using the Ballesteros-Weinstein (A), Wootten (B), Pin (C) and Wang (F) schemes in the format X.50 where X is the helix (1–7) followed by the residue position relative to the most conserved (50). Where GPCR crystal structure superposition shows a numbering offset due to helical bulges or constrictions, the structure-based GPCRDB number can be appended separated by an x, for example 5.42x43, or used as a substitute, for example 5x43. In a similar way, it is possible to combine the equivalent numbers for the different GPCR classes after appending the class name, for example 3.43a.47b. Generic numbers can also be combined with receptor-specific numbers by using superscript, for example S348^{5.47x48} or S5.47x48³⁴⁸. Generic residue numbers can be assigned to any sequence, structure or pdf article at GPCRDB.

numbering as new crystal structure templates become available. In fact, there are seven alignments – one for each helix. Most receptors and helices have a sufficiently high overall sequence similarity, and the bulk of the others can be anchored on highly conserved residue positions (e.g., the X.50 references). The few remaining are mostly orphan receptors with atypical sequences [75].

Upcoming crystal structures are expected to reveal additional offsets of the sequence-based numbering, while improving the structure-based system. The flip-side of the coin is that this requires the GPCRDB numbers to be updated for the given subset of receptors and offset residue positions. The GPCRDB numbering scheme has been designed so that the numbers are adjusted only when a newly determined GPCR structure reveals a new bulge or constriction that has not been predicted by homology. It should be pointed out that all abundant offsets, affecting many receptors, have most likely already been covered. Firstly, all bulges (Figure 3A–I) have a proline residue 3–5 positions downstream, and in the largest class (A) there are no more highly conserved proline residues in the 7TM that allow additional frequent bulges. Secondly, most other bulges and constrictions have so far only been observed in a single crystal structure and have a low expected frequency, indicating that future adjustments will be limited to small receptor subfamilies or a subset of their members. Consequently, determination of structures covering novel receptor families will be the most valuable.

Conformational plasticity and flexibility are intrinsic features of individual receptors that have to be accounted for in any analysis and generalization of structural information for GPCRs. So far, generic structure-based numbering has shown good tolerance to conformational changes in 7TM region. Thus, our analysis shows that the bulges and constrictions do not change between multiple crystal structures of the same GPCR, when available, even though these structures were determined using different complexes, resolutions, stabilization constructs, or crystal packing forms. Even when relatively large conformational changes between different activation states are considered, the corresponding helical distortions do not involve formation of a full bulge/constriction. This does not preclude a possibility of such changes being found in some of the newly solved GPCR structures in the future – in this case careful

analysis should be applied to understand the mechanism and exclude potential artifacts.

Concluding remarks

The first crystal structures of the class B, C, and F GPCRs have opened up the field for receptor function studies and drug design. In addition, cross-class sequence alignments can now be constructed, enabling us to uncover the common cogs and cranks within the 7TM machinery. We have described the schemes for generic numbering of such residue hotspots for chemical structure-activity relationships, pharmacological effects of receptor mutants, and structural mechanisms. Furthermore, the multiplicity of structures has uncovered helix bulges and constrictions, making it clear that the correct mapping of equivalent residues requires the gapping of sequence alignments based on crystallographic evidence. Herein, GPCRDB presents such reference structure-based alignments and residue numbers. It is recommended that a combination of sequence-based and structure-based GPCRDB numbers is used, for example 2.58x57, unless the comparison only includes crystallized receptors.

Acceptance by the GPCR community will initially be determined by simplicity – in other words counting numbers within the specific sequence (alignment) as compared to looking it up in the GPCRDB alignments or tables. In the longer term, it is more crucial how many offsets accumulate in the sequence-based numbering and how many can be accepted while still preserving the utility of the system such that the same number actually refers to the equivalent residues across different GPCRs. The GPCRDB numbering system has been fully endorsed by the International Union of Basic and Clinical Pharmacology Committee on Receptor Nomenclature and Drug Classification (NC-IUPHAR). NC-IUPHAR, together with its database GuideToPharmacology [76], will be working closely with GPCRDB to encourage acceptance of this system by the wider GPCR community.

Acknowledgments

The Lundbeck Foundation provided financial support to V.I. and D.E.G., who was also funded by the Carlsberg Foundation. V.C., V.K., and R.C.S. are supported by NIH grants U54 GM094618 and P01 DA035764. G.V. acknowledges TIPharma for financial supports. This work was also supported by COST Action CM1207 (GLISTEN). Kasper Harpsøe is acknowledged for discussions on structure superposition and residue numbering.

References

- 1 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
- 2 Pierce, K.L. *et al.* (2002) Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* 3, 639–650
- 3 Overington, J.P. *et al.* (2006) How many drug targets are there? *Nat. Rev. Drug Discov.* 5, 993–996
- 4 Kolakowski, L.F., Jr (1994) GCRDB: a G-protein-coupled receptor database. *Receptors Channels* 2, 1–7
- 5 Fredriksson, J.P. *et al.* (2003) The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* 63, 1256–1272
- 6 Venkatakrishnan, A.J. *et al.* (2013) Molecular signatures of G-protein-coupled receptors. *Nature* 494, 185–194
- 7 Tehan, B.G. *et al.* (2014) Unifying family A GPCR theories of activation. *Pharmacol. Ther.* 143, 51–60

- 8 Katritch, V. *et al.* (2013) Structure-function of the G protein-coupled receptor superfamily. *Annu. Rev. Pharmacol. Toxicol.* 53, 531–556
- 9 Lagerstrom, M.C. and Schioth, H.B. (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. *Nat. Rev. Drug Discov.* 7, 339–357
- 10 Gloriam, D.E. (2013) Chemogenomics of allosteric binding sites in GPCRs. *Drug Discov. Today Technol.* 10, e307–e313
- 11 Garland, S.L. (2013) Are GPCRs Still a Source of New Targets? *J. Biomol. Screen.* 18, 947–966
- 12 Congreve, M. *et al.* (2011) Progress in structure based drug design for G protein-coupled receptors. *J. Med. Chem.* 54, 4283–4311
- 13 Gloriam, D.E. *et al.* (2009) Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design. *J. Med. Chem.* 52, 4429–4442
- 14 Ballesteros, J.A. and Weinstein, H. (1995) Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* 25, 366–428
- 15 Isberg, V. *et al.* (2014) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.* 42, D422–D425
- 16 Oliveira, L. *et al.* (1993) A common motif in G-protein-coupled seven transmembrane helix receptors. *J. Comput. Aided Mol. Des.* 7, 649–658
- 17 Baldwin, J.M. *et al.* (1997) An alpha-carbon template for the transmembrane helices in the rhodopsin family of G-protein-coupled receptors. *J. Mol. Biol.* 272, 144–164
- 18 Baldwin, J.M. (1993) The probable arrangement of the helices in G protein-coupled receptors. *EMBO J.* 12, 1693–1703
- 19 Schwartz, T.W. *et al.* (1995) Molecular mechanism of action of non-peptide ligands for peptide receptors. *Curr. Pharm. Des.* 1, 325–342
- 20 Schwartz, T.W. (1994) Locating ligand-binding sites in 7TM receptors by protein engineering. *Curr. Opin. Biotechnol.* 5, 434–444
- 21 Wooten, D. *et al.* (2013) Polar transmembrane interactions drive formation of ligand-specific and signal pathway-biased family B G protein-coupled receptor conformations. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5211–5216
- 22 Siu, F.Y. *et al.* (2013) Structure of the human glucagon class B G-protein-coupled receptor. *Nature* 499, 444–449
- 23 Hollenstein, K. *et al.* (2013) Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* 499, 438–443
- 24 Pin, J.P. *et al.* (2003) Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacol. Ther.* 98, 325–354
- 25 Dore, A.S. *et al.* (2014) Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain. *Nature* 511, 557–562
- 26 Wang, C. *et al.* (2014) Structural basis for Smoothed receptor modulation and chemoresistance. *Nat. Commun.* 5, 4355
- 27 Hollenstein, K. *et al.* (2014) Insights into the structure of class B GPCRs. *Trends Pharmacol. Sci.* 35, 12–22
- 28 de Graaf, C. *et al.* (2011) Structure-based discovery of allosteric modulators of two related class B G-protein-coupled receptors. *ChemMedChem* 6, 2159–2169
- 29 Molck, C. *et al.* (2012) Pharmacological Characterization and Modeling of the Binding Sites of Novel 1,3-Bis(pyridinylethynyl)benzenes as Metabotropic Glutamate Receptor 5-Selective Negative Allosteric Modulators. *Mol. Pharmacol.* 82, 929–937
- 30 Gloriam, D.E. *et al.* (2011) Chemogenomic Discovery of Allosteric Antagonists at the GPRC6A Receptor. *Chem. Biol.* 18, 1489–1498
- 31 Wu, H. *et al.* (2014) Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science* 344, 58–64
- 32 Wang, C. *et al.* (2013) Structure of the human smoothed receptor bound to an antitumour agent. *Nature* 497, 338–343
- 33 Bortolato, A. *et al.* (2014) Structure of Class B GPCRs: New Horizons for Drug Discovery. *Br. J. Pharmacol.* 171, 3132–3145
- 34 Oliveira, L. *et al.* (1999) A low resolution model for the interaction of G proteins with G protein-coupled receptors. *Protein Eng.* 12, 1087–1095
- 35 Rasmussen, S.G. *et al.* (2011) Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* 477, 549–555
- 36 Shi, L. and Javitch, J.A. (2002) The binding site of aminergic G protein-coupled receptors: the transmembrane segments and second extracellular loop. *Annu. Rev. Pharmacol. Toxicol.* 42, 437–467
- 37 Kooistra, A.J. *et al.* (2013) A structural chemogenomics analysis of aminergic GPCRs: lessons for histamine receptor ligand design. *Br. J. Pharmacol.* 170, 101–126
- 38 Scholten, D.J. *et al.* (2012) Pharmacological modulation of chemokine receptor function. *Br. J. Pharmacol.* 165, 1617–1643
- 39 Gao, Z.G. *et al.* (2002) Identification by site-directed mutagenesis of residues involved in ligand recognition and activation of the human A3 adenosine receptor. *J. Biol. Chem.* 277, 19056–19063
- 40 Parrill, A.L. and Tigyi, G. (2013) Integrating the puzzle pieces: the current atomistic picture of phospholipid-G protein coupled receptor interactions. *Biochim. Biophys. Acta* 1831, 2–12
- 41 Malherbe, P. *et al.* (2003) Mutational analysis and molecular modeling of the allosteric binding site of a novel, selective, noncompetitive antagonist of the metabotropic glutamate 1 receptor. *J. Biol. Chem.* 278, 8340–8347
- 42 Malherbe, P. *et al.* (2006) Comparison of the binding pockets of two chemically unrelated allosteric antagonists of the mGlu5 receptor and identification of crucial residues involved in the inverse agonism of MPEP. *J. Neurochem.* 98, 601–615
- 43 Molck, C. *et al.* (2014) mGluR5: Exploration of Orthosteric and Allosteric Ligand Binding Pockets and Their Applications to Drug Discovery. *Neurochem. Res.*
- 44 Petrel, C. *et al.* (2004) Positive and negative allosteric modulators of the Ca²⁺-sensing receptor interact within overlapping but not identical binding sites in the transmembrane domain. *J. Biol. Chem.* 279, 18990–18997
- 45 Miedlich, S.U. *et al.* (2004) Homology modeling of the transmembrane domain of the human calcium sensing receptor and localization of an allosteric binding site. *J. Biol. Chem.* 279, 7254–7263
- 46 Mirzadegan, T. *et al.* (2003) Sequence analyses of G-protein-coupled receptors: similarities to rhodopsin. *Biochemistry* 42, 2759–2767
- 47 de Graaf, C. *et al.* (2008) Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. *Proteins* 71, 599–620
- 48 Kufareva, I. *et al.* (2014) Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure*
- 49 Kufareva, I. *et al.* (2011) Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. *Structure* 19, 1108–1126
- 50 Michino, M. *et al.* (2009) Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discov.* 8, 455–463
- 51 Yohannan, S. *et al.* (2004) The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* 101, 959–963
- 52 van der Kant, R. and Vriend, G. (2014) Alpha-bulges in G protein-coupled receptors. *Int. J. Mol. Sci.* 15, 7841–7864
- 53 Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8, 52–56
- 54 Govaerts, C. *et al.* (2001) The TXP motif in the second transmembrane helix of CCR5. A structural determinant of chemokine-induced activation. *J. Biol. Chem.* 276, 13217–13225
- 55 Minami, M. *et al.* (1996) A single residue, Lys108, of the delta-opioid receptor prevents the mu-opioid-selective ligand [D-Ala2,N-MePhe4,Gly-ol5]enkephalin from binding to the delta-opioid receptor. *Mol Pharmacol* 50, 1413–1422
- 56 Wu, B. *et al.* (2010) Structures of the CXCR4 Chemokine GPCR with Small-Molecule and Cyclic Peptide Antagonists. *Science* 330, 1066–1071
- 57 Roumen, L. *et al.* (2011) In Silico Veritas: The Pitfalls and Challenges of Predicting GPCR-Ligand Interactions. *Pharmaceuticals* 4, 1196–1215
- 58 Wu, H. *et al.* (2012) Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* 485, 327–332
- 59 Thompson, A.A. *et al.* (2012) Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485, 395–399
- 60 Fukuda, K. *et al.* (1995) Identification of the amino acid residues involved in selective agonist binding in the first extracellular loop of the delta- and mu-opioid receptors. *FEBS Lett.* 373, 177–181
- 61 Warne, T. *et al.* (2011) The structural basis for agonist and partial agonist action on a beta(1)-adrenergic receptor. *Nature* 469, 241–244
- 62 Warne, T. *et al.* (2012) Crystal structures of a stabilized beta1-adrenoceptor bound to the biased agonists bucindolol and carvedilol. *Structure* 20, 841–849

- 63 Rosenbaum, D.M. *et al.* (2011) Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. *Nature* 469, 236–240
- 64 Shimamura, T. *et al.* (2011) Structure of the human histamine H1 receptor complex with doxepin. *Nature* 475, 65–70
- 65 Wieland, K. *et al.* (1999) Mutational analysis of the antagonist-binding site of the histamine H(1) receptor. *J. Biol. Chem.* 274, 29994–30000
- 66 Haga, K. *et al.* (2012) Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* 482, 547–551
- 67 Kruse, A.C. *et al.* (2012) Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* 482, 552–556
- 68 Heitz, F. *et al.* (1999) Site-directed mutagenesis of the putative human muscarinic M2 receptor binding site. *Eur. J. Pharmacol.* 380, 183–195
- 69 Istyastono, E.P. *et al.* (2011) Molecular determinants of ligand binding modes in the histamine H(4) receptor: linking ligand-based three-dimensional quantitative structure-activity relationship (3D-QSAR) models to in silico guided receptor mutagenesis studies. *J. Med. Chem.* 54, 8136–8147
- 70 Lim, H.D. *et al.* (2010) Molecular determinants of ligand binding to H4R species variants. *Mol. Pharmacol.* 77, 734–743
- 71 Horn, F. *et al.* (2003) GPCRDB information system for G protein-coupled receptors. *Nucleic Acids Res.* 31, 294–297
- 72 Horn, F. *et al.* (1998) GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.* 26, 275–279
- 73 Vroling, B. *et al.* (2011) GPCRDB: information system for G protein-coupled receptors. *Nucleic Acids Res.* 39, 309–319
- 74 Vroling, B. *et al.* (2011) Integrating GPCR-specific information with full text articles. *BMC Bioinformatics* 12, 362
- 75 Gloriam, D.E. *et al.* (2007) The G protein-coupled receptor subset of the rat genome. *BMC Genomics* 8, 338
- 76 Pawson, A.J. *et al.* (2014) The IUPHAR/BPS Guide to PHARMACOLOGY: an expert-driven knowledgebase of drug targets and their ligands. *Nucleic Acids Res.* 42, D1098–D1106
- 77 Olender, T. *et al.* (2013) HORDE: comprehensive resource for olfactory receptor genomics. *Methods Mol. Biol.* 1003, 23–38
- 78 Palczewski, K. *et al.* (2000) Crystal structure of rhodopsin: A G protein-coupled receptor. *Science* 289, 739–745
- 79 Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242